CONFIDENTIAL

15 July 1964

Statement of Initial OBJECTIVES

CODIB

Ad Hoc Problem Group VI

Research and Development

I. Problem Group Objective

To develop a Community coordinated Research and Development Program in the general field of non-numerical data processing, including automatic input devices, associative memories, machine translation and other specific EDP areas for application in intelligence information processing operations.

II. Initial Problem Parameters

A. Scope

1. The program is to include both internal and sponsored external Research and Development on systems, techniques, and equipment with potential to improve information processing operations in the Intelligence Community.

2. Information processing, on all forms, sources, subjects, and languages of recognized importance to intelligence to be taken into consideration.

3. Both hardware and software fall within scope of the program including indexing techniques, file organization, intelligence oriented machine programming, non-machine communication, communication degarbling techniques, automatic abstracting, file conversion, file organization, trends analysis, language translation, content addressing, and associated memories as well as electronic display equipment remote input/output stations, character sensing devices, etc.

CONFIDENTIAL

-2-

4. Techniques and means to facilitate the training of
    Community personnel in the newly developed equipment and
    techniques to be considered within the scope of the
    program.

5. Techniques and means for the introduction and testing
    of Research and Development results in the Community
    are within the scope of the program.

6. The controlled and qualified monitorship of sponsored
    external Research and Development is to be an aspect of
    the Community program.

B. Characteristics of the Coordinated R & D Program

1. Program will take into account the full range of
    intelligence interests and needs.

2. Program will provide for indications of R & D priorities within
    and between projects.

3. Program will provide for reporting to the Community
    of overlap as well as gaps in the sponsored research and
    development.

4. Pertinent details of R & D sponsored outside the Community
    will be included in the reporting system.

5. The program will include provisions for means to determine
    community needs for particular R & D efforts through
    specification of information processing problems, and
    mechanisms for feeding these specs into the R & D
    Community.

6. Projects should be related in some perspective to the
   sponsor and *prospective* results should be related to Community

   needs and interest.

7. One of the principal features of the program will be

   its own updating and reporting for current awareness.

C. Constraints

1. Only R & D of recognized intelligence interest to be

   considered.

2. Plan must account for present Departmental R & D projects

   and priorities.

3. Security classifications assigned to projects by

   sponsoring organizations will have to be observed.

4. Full advantage must be taken of non-Community sponsored

   R & D.

5. Relationship of reporting system to other R & D reporting

   systems (e.g., DDC, NSF, SIE, etc.) must be specified to

   avoid undesired repetition.

-4-

III. Initial Problem Group Tasks

1. Obtain descriptions of all in-house and external R & D

   activities and projects. Descriptions to contain roughly

   the same elements as are on the Scientific Information

   Exchange (SIE) form ADP/RD-1. Descriptions should carry

   ████████████████ whatever security classification is required

   to fully describe the project and the particular intelligence

   processing needs addressed by the efforts.

2. Categorize and summarize present activities and projects by

   area of processing interest.

3. Solicit statements of R & D needs from each agency and consolidate.

4. Compare needs with present efforts for balance between problem

   areas.

5. Develop alternative rough configurations of a coordinated

   R & D program and contrast them with the aggregate of present

   departmental programs.

6. Postulate procedures and mechanisms for the detailed development

   and maintenance of a community R & D program and estimate

   required efforts.

7. Determine additional tasks required of the problem group.

8. Report to CODIB.

ILLEGIB

-5-

IV. Schedule

1. Problem group members to be designated by _____

2. Initial report to CODIB to be ready by _____

V. Membership

Participants will require full security clearances and be

knowledgeable of the state of the art in IP and be up to date on

research and development in the field, particulalry R & D under way

██████ within and sponsored by USIB Agencies.

VI. Guidance

The problem groups will be under guidance of CODIB.

... Comments elaborate to furnish reference
for R+D Task Team.

It would appear advisable that
for the "Identification and
understanding of problem areas" (III A)
that the R+D group look to the
other ad groups for initial problem
specification in their areas of
concern. This would minimize
the tendency for R+D to overlap
or run ahead of the other ad
hoc groups in their assigned
areas. There will be some overlap between
most of the ad hoc working groups

however, R+D will address problems in _all_ of the other group areas. In fact, R+D may represent the solution to many problems identified and defined by the other ad hoc groups.

Therefore a parallel approach represented by the problems assigned the various ad hoc groups must be closely coordinated or we will fund our groups as entering the

organizations were prudently, asking the same or similar questions as they proceed giving rise to the allegation that our *collective* effort is neither well planned or coordinated and extremely expensive both from our standpoint and our impact on the operating environment.

It might be well to reconsider the "need" orientation of the R+D effort. We still have a need to know who's doing

what in R+D that may be
pertinent to intell. problems when
we (the ad hoc groups) have
specified them . State of the
art in print readers, auto-
indexing, auto abstracting,
machine translation, trend
analysis, pattern recognition
are still fruitful areas in
which we can stipulate
a requirement from the outset.

CODIB Task Team VI
Application Area: Language Processing

Description and Scope:

By agreement this area of activity includes abstracting,
indexing, extracting, and formatting functions. It does not include
"machine translation" which is regarded as a separate area.

In explanation of the functions of interest it should be
noted that abstracting and indexing require a high degree of human
intellectual effort. Extracting is largely a clerical operation and
follows specific rules. Formatting refers to the rearrangement
of data according to specified patterns and is largely a machine
operation.

Discussion of Problem Areas:

With reference to the kinds of problems which are involved
in the various functions indicated above, the following distinctions
can be made:

1. Problems of document generation.
    (related to "abstracting")

2. Problems concerned with the design and
detailing of clerical and machine procedures.
    (related to "extracting, " "formatting")

3. Problems of description or representation
of natural language texts.
    (related to "indexing")

Of the problem areas indicated, our estimate is that the area
concerned with the representation of natural language texts is by far
the most important one at this time and should receive particular
attention in the further work of this committee.

Why is it important to conduct R&D on "representation of natural language texts"?

1. The problem area is a basic one. It deals with the nature of what flows in systems of interest. What are the units? Are they discrete? Are they complexes?

2. Evidences are building up that units used in present systems are not acceptable. Representations are too coarse; they do not have enough fine structure and do not provide enough dis- criminating power.

3. Any new insight in this area can have direct and profound effect on the design of systems and hardware. A strong case can be made that many system functions which are now in use are required only because of the deficiencies of the units now employed in the system.

4. Progress has been made in many of the disciplines which have a bearing on the problem. Work in linguistics, logic, semantic theory, computer processing, has shown pertinent and significant results in the last few years. These results should be coordinated and evaluated with respect to the representation problem.

January 6, 1965

25X1A